

Kohler, Britta

Zur Rezeption von TIMSS durch Lehrerinnen und Lehrer

Unterrichtswissenschaft 30 (2002) 2, S. 158-188



Quellenangabe/ Reference:

Kohler, Britta: Zur Rezeption von TIMSS durch Lehrerinnen und Lehrer - In: Unterrichtswissenschaft 30 (2002) 2, S. 158-188 - URN: urn:nbn:de:0111-opus-76844 - DOI: 10.25656/01:7684

<https://nbn-resolving.org/urn:nbn:de:0111-opus-76844>

<https://doi.org/10.25656/01:7684>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Unterrichtswissenschaft

Zeitschrift für Lernforschung

30. Jahrgang / 2002 / Heft 1

Thema:

7a+b, 12c

PISA – Konzept und Ergebnisse

Verantwortliche Herausgeber:

Jürgen Baumert, Manfred Prenzel

Jürgen Baumert, Manfred Prenzel:

Einführung

98

Michael Neubrand, Eckhard Klieme, Oliver Lüdtke, Johanna Neubrand:

Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test
zur mathematischen Grundbildung

100X

DLG ✓

Manfred Prenzel, Peter Häußler, Jürgen Rost, Martin Senkbeil:

Der PISA-Naturwissenschaftstest:

Lassen sich die Aufgabenschwierigkeiten vorhersagen?

120

Joachim Wirth, Eckhard Klieme:

Computer literacy im Vergleich zwischen Nationen,

Schulformen und Geschlechtern

136X

HP ✓

Allgemeiner Teil

Britta Kohler:

Zur Rezeption von TIMSS durch Lehrerinnen und Lehrer

158

Hinweise für die Autoren

190

97

Britta Kohler

Zur Rezeption von TIMSS durch Lehrerinnen und Lehrer

The Reception of TIMSS by Teachers

Über die Rezeption leistungsvergleichender Studien durch Lehrerinnen und Lehrer liegen bislang kaum Daten vor. Auch ist noch weitgehend ungeklärt, in welcher Weise beteiligten Lehrkräften die Ergebnisse dieser Untersuchungen rückgemeldet werden sollen. Mit Hilfe einer Fragebogenstudie sollte deshalb unter anderem der Frage nachgegangen werden, wie Lehrkräfte die TIMS-Studie rezipierten und wie sie generell Vorhaben externer Evaluation gegenüberstehen. An der im Jahr 2000 durchgeführten zweiteiligen Fragebogenstudie mit integrierter Informationseinheit nahmen 201 Lehrerinnen und Lehrer allgemeinbildender Schulen aus zwei Oberschulamtsbezirken Baden-Württembergs teil. Es zeigte sich, dass die Lehrkräfte über die TIMS-Studie nur wenig informiert waren und sie die Ergebnisse für nur mäßig bedeutsam hielten. Bei der Suche nach möglichen Erklärungen für das nur mittelmäßige Abschneiden der deutschen Schülerinnen und Schüler zeigten sich Attributionsvoreingenommenheiten in Form von externalen Ursachenzuschreibungen. Die Lehrkräfte gaben sich leistungsvergleichenden Studien gegenüber aufgeschlossen und konnten sich eine Teilnahme unter der Bedingung von Anonymität überwiegend vorstellen.

To date there is comparatively little data existing about the reception of performance comparison studies by educators. It is also unclear how the participant educators should be informed about the results of said studies. A questionnaire was conducted to determine how educators received the results of the TIMS-Study and what their general impression of external evaluation is. The results of a two part questionnaire, including an information unit, conducted in 2000 amongst 201 educators from two school districts in Baden-Württemberg indicated that the educators were little informed about the TIMS-Study and that they hold the results to be of middling importance. The search for possible explanations of the mediocre performance of German students indicated that educators had an attribution bias towards external factors. The educators had a positive disposition toward comparative performance studies and predominantly they could imagine to participate so long as they remain anonymous.

1. Einleitung und Fragestellungen

Die TIMS-Studie (Baumert, Lehmann, Lehrke, Schmitz, Clausen, Hosenfeld, Köller & Neubrand, 1997) löste in der Fachwelt und auch in der interessierten Öffentlichkeit bekanntlich eine Reihe unterschiedlicher, zum Teil heftiger Reaktionen aus; und so war in Anlehnung an den Sputnik-Schock der 60er Jahre bald vom „TIMSS-Schock“ die Rede. Die Gegner vergleichender Leistungsstudien sahen eine massive Verstärkung der Leistungsorientie-

rung, eine Hinwendung zu einem „teaching to the test“, eine verstärkte Kontrolle des Schulwesens, die Verschwendung von Ressourcen, eine Demotivation von Lehrenden und Lernenden mit ungünstigen Kontextbedingungen, Testkosmetik an den Schulen, zunehmendes Scheinwissen und in der Folge gar die „Verbote eines internationalen Schulmodells ... [auf uns zukommen], das nach dem Prinzip McSchool von kommerziell wirtschaftenden Bildungsketten betrieben wird“ (Demmer, 2000, S. 34). Die Befürworter hingegen sprachen von der „wohl ambitioniertesten internationalen Schulleistungsuntersuchung, die zum Ende des vergangenen Jahrzehnts durchgeführt wurde“ (Bos & Postlethwaite, 2001, 253), die einen „Glücksfall“ (Blum, 2000, 26) darstelle, einen Prozess der Qualitätsentwicklung in Gang bringen konnte und die öffentliche und professionelle Aufmerksamkeit neu zu strukturieren vermochte (Baumert, Bos & Lehmann, 2000). Auch wurde von der Pro-Seite immer wieder betont, dass es das Verdienst der TIMS-Studie sei, anderen, bereits vorhandenen Studien zur notwendigen Aufmerksamkeit verhelfen zu haben (Sygusch, 1999). Und schließlich wurde darauf hingewiesen, dass die TIMS-Studie ungezählte Initiativen, Reform- und Forschungsvorhaben auf den Weg bringen (Weinert, 2001b) und letztlich eine für überfällig gehaltene empirische Wende in Bildungspolitik und Erziehungswissenschaft einleiten konnte.

Schon kurze Zeit nach der Veröffentlichung der TIMS-Studie, also zu einem Zeitpunkt, zu dem die Diskussion noch ganz am Anfang stand, verständigten sich Bildungspolitik, Schulaufsicht und an der Thematik interessierte Erziehungswissenschaftlerinnen und Erziehungswissenschaftler sowie Kolleginnen und Kollegen anderer Disziplinen darauf, in Zukunft verstärkt auf vergleichende Schulleistungsstudien zu setzen. Getragen von der Überzeugung, dass externe Evaluation notwendiges Orientierungswissen bereitstellen und schulische bzw. unterrichtliche Qualitätsentwicklung voranbringen kann, hoben sie mehrere Leistungsstudien aus der Taufe bzw. beschlossen die Beteiligung an internationalen Studien sowohl für die Sekundarstufe (z. B. PISA) als auch für die Primarstufe (IGLU als deutsche Teilstudie der IEA-PIRLS). Ein Punkt, der dabei regelmäßig diskutiert wurde, war die Frage der Rückmeldung der Ergebnisse an die beteiligten Lehrkräfte und überhaupt der Einbezug von Lehrerinnen und Lehrern in die Konzeption der jeweiligen Studie. Ganz offensichtlich war man zu der Überzeugung gelangt, dass eine Schulleistungsstudie, deren Vorhandensein und deren Ergebnisse Lehrkräfte zunächst mehr oder weniger zufällig aus der Tagespresse erfahren, nicht unbedingt geeignet ist, Qualitätsentwicklung in Gang zu setzen. Und so war insbesondere bei den Vollerhebungen LAU (Hamburg) und MARKUS (Rheinland-Pfalz), aber auch bei der stichprobenartig angelegten Untersuchung QuaSUM (Brandenburg) bereits von Anfang an eingeplant, den beteiligten Lehrkräften möglichst detaillierte und für sie relevante, klassenbezogene Rückmeldungen zu geben, die teilweise auch an Eltern und Schülerinnen und Schüler weitergegeben werden können: „Der wissenschaftliche Bericht (...) und die klassenbezogenen Rückmeldungen sollen Anlass sein, unter bestimmten Fragestellungen die pädagogische und unter-

richtliche Arbeit zu reflektieren und sie gegebenenfalls weiterzuentwickeln oder die neuen Erkenntnisse in die Schulprogrammgestaltung und Schulentwicklung einfließen zu lassen“ (Behörde für Schule, Jugend und Berufsbildung Hamburg, 2000. S. 1).

Doch während Einigkeit darüber besteht, dass die an Schulleistungstudien beteiligten Lehrkräfte über die Untersuchungsergebnisse informiert werden sollten - sei es aus Gründen der Fairness oder als Anstoß zur Qualitätsentwicklung -, ist noch weitgehend ungeklärt, in welcher Weise die Rückmeldungen zu erfolgen haben (Helmke & Schrader, 2001; Klieme, Baumert & Schwippert, 2000). Die Unklarheiten beginnen dabei vergleichsweise früh, nämlich schon bei der Frage, welche Daten überhaupt sinnvollerweise rückgemeldet werden sollen (Rolff, 2001). Dabei ist eines unübersehbar: Die mit hohem Aufwand gewonnenen Ergebnisse vergleichender Leistungsstudien können nur dann Veränderungen unmittelbarer Art in der schulischen Arbeit einleiten, wenn sie von den Lehrerinnen und Lehrern tatsächlich wahrgenommen, adäquat erfasst und überdies für bedeutsam und zutreffend gehalten werden. Doch um eine solche Rezeption bewirken zu können, müssen Anlage und Ergebnisse der jeweiligen Untersuchung in geeigneter Weise kommuniziert werden, d. h. in einer Weise, die weder zu Desinteresse führt noch Entmutigung oder Reaktanz erzeugt.

Das Ziel der nachfolgend beschriebenen Untersuchung war es deshalb herauszufinden, in welcher Weise die TIMS-Studie durch Lehrerinnen und Lehrer rezipiert worden war. Auf der Grundlage der gewonnenen Erkenntnisse werden dann erste Überlegungen bezüglich der Frage angestellt, was bei der Konzeption externer Evaluationsstudien und bei der Kommunikation ihrer Ergebnisse an Lehrkräfte zu beachten ist, damit externe Evaluation tatsächlich einen Beitrag zur unterrichtlichen Qualitätsentwicklung leisten kann.

Im einzelnen wurde folgenden Fragestellungen nachgegangen:

- (1) Inwieweit ist den Lehrerinnen und Lehrern die TIMS-Studie bekannt?
- (2) Wie schätzen die Lehrerinnen und Lehrer die Ergebnisse der TIMS-Studie ein?
- (3) Wie stehen die Lehrenden generell zu (kultur-)vergleichenden Leistungsstudien?
- (4) Worauf führen die Lehrerinnen und Lehrer das nur mittelmäßige Abschneiden der deutschen Schülerinnen und Schüler zurück? Ergeben sich Hinweise auf Attributionsvoreingenommenheiten (Heckhausen, 1989; Weiner, 1994)?
- (5) Wird von den befragten Lehrkräften ein Bezug von den Ergebnissen zur eigenen Situation hergestellt bzw. in welcher Weise erfolgt dieser Bezug?
- (6) Führt und führt TIMSS bei den Lehrenden zu dem Wunsch und der Bereitschaft, den eigenen Unterricht zu verändern und zu verbessern, war

und ist die TIMS-Studie somit auf diesem direkten Wege in der Lage, Qualitätsverbesserung anzustoßen bzw. voranzubringen?

- (7) Inwieweit wären die Lehrerinnen und Lehrer bereit, selbst an einer entsprechenden Studie teilzunehmen?

2. Methode

2.1 Stichprobe

An der Untersuchung nahmen insgesamt 201 Lehrerinnen und Lehrer allgemeinbildender Schulen aus zwei Oberschulamtsbezirken Baden-Württembergs teil. Sie verteilten sich weitgehend gleichmäßig über die verschiedenen Schularten: 61 von ihnen unterrichteten an Grund- und Hauptschulen (vier Schulen), 53 an Realschulen (drei Schulen) und 87 an Gymnasien (sechs Schulen). Während die Befragungen der Lehrkräfte an Grund- und Hauptschulen sowie an Realschulen grundsätzlich an Gesamtlehrerkonferenzen vorgenommen wurden, hier also immer vollständige Kollegien befragt werden konnten, geschah die Befragung der Gymnasiallehrkräfte zum Teil auch an Fachkonferenzen mit mehr oder weniger ausgeprägtem Freiwilligkeitscharakter. Eine positive Verzerrung der Untersuchungsergebnisse bei den Gymnasiallehrkräften kann somit nicht ausgeschlossen werden. Gewonnen wurden die Befragten über die jeweiligen Schulleitungen, die sich mehrheitlich selbst an der Untersuchung beteiligten. Bei den Grund- und Hauptschulen wirkte auch ein Schulamt unterstützend mit. Nach der Kontaktaufnahme war etwa jede zweite Schule bzw. Schulleitung bereit, sich an der Untersuchung zu beteiligen. Das mit der Anfrage immer gleichzeitig formulierte Angebot, über TIMSS informiert zu werden und eine Diskussion im Kollegium führen zu können, wurde sehr unterschiedlich aufgenommen und umgesetzt.

Von den befragten Lehrerinnen und Lehrern waren 86 weiblichen und 113 männlichen Geschlechts. Das durchschnittliche Alter betrug knapp 47 Jahre, lag damit also geringfügig unter dem Landesdurchschnitt. Jeweils ca. 27 % der Lehrkräfte unterrichteten nur Mathematik und/oder Naturwissenschaften bzw. weder Mathematik noch Naturwissenschaften. Der Anteil der Lehrenden, die nur Mathematik und/oder Naturwissenschaften unterrichteten, war aufgrund der unterschiedlichen Lehrauftragsverteilungen an den verschiedenen Schularten und aufgrund der Art der Stichprobenziehung unter den Gymnasiallehrkräften am höchsten (ca. 50 %). Genau 60 % der Befragten verfügten über eine Zusatzqualifikation und/oder Zusatzfunktion, waren also z. B. Mentor, Beratungslehrer oder Schulleiterin. 24 % von ihnen hatten keine Kinder, während 71 % der befragten Lehrkräfte Eltern schulpflichtiger oder ehemals schulpflichtiger Kinder waren.

Die Befragung erfolgte in anonymer Weise. Da aber mit generierbaren Codes gearbeitet wurde, konnte die Möglichkeit einer Nachbefragung offen gehalten werden.

2.2 Design und Ablauf der Untersuchung

Die Untersuchung wurde als zweiteilige Fragebogenstudie mit integrierter Informationseinheit konzipiert. Sie diente vorwiegend der Erhebung quantitativer Daten, sollte darüber hinaus aber auch qualitative Daten erbringen. Durchgeführt wurde sie im Jahr 2000. Die Erhebung fand grundsätzlich im Rahmen von (Gesamtlehrer-)konferenzen statt und wurde wie folgt realisiert: Zunächst wurden die Anwesenden, mehr formal als inhaltlich, über die Untersuchung informiert und mit der Kennzeichnung der Bögen durch generierbare Codes vertraut gemacht. Anschließend folgte der erste Teil der Untersuchung, in dem vor allem personenbezogene Daten und das bereits vorhandene Wissen über die TIMS-Studie abgefragt wurden. Die dann eingeschobene, standardisierte Informationseinheit von ca. 10 Minuten Dauer diente der Herstellung einer für den zweiten Teil der Befragung ausreichenden Wissensbasis für alle. Es wurde dabei in aller Kürze vermittelt, wofür die Abkürzung TIMSS steht, wer für die Durchführung der Studie zuständig gewesen war, aus welchen Bereichen Aufgaben gestellt worden waren und wie die deutschen Schülerinnen und Schüler im Verhältnis zu den Lernenden anderer Staaten abgeschnitten hatten. Die besonderen Schwächen der deutschen Schülerinnen und Schüler in den Bereichen Anwenden und Problemlösen wurden hierbei kurz erwähnt. Wertungen und eigene Einschätzungen wurden vermieden. Rückfragen wurden in der Regel nicht sofort, sondern erst nach dem zweiten Befragungsteil beantwortet. In diesem sollten die Lehrerinnen und Lehrer Stellung zu den Ergebnissen der TIMS-Studie beziehen und auch angeben, welche Konsequenzen sie daraus für sich und ihren Unterricht ziehen. Im Anschluss an die Befragung fand dann - je nach Interesse und verfügbarer Zeit - noch eine Diskussion im Kollegium statt. In diesem Kontext wurde zumeist ein Interesse an den Resultaten der Untersuchung bekundet.

2.3 Instrumente

Die erste Hälfte des aus zwei Teilen bestehenden Fragebogens diente im Wesentlichen der Erfassung personenbezogener Daten sowie des subjektiven Informationsstandes und des vorhandenen Wissens bezüglich der TIMS-Studie. Während zur Erhebung des Vorwissens vier offen zu beantwortende Fragen formuliert wurden, stehen für die Erfassung personenbezogener Daten und des subjektiven Informationsstandes insgesamt 14 Fragen mit Antwortvorgaben (z. B. Fächerkombination, Zusatzqualifikationen) oder Rating-Skalen (Belastung, Schulklima) mit vier Stufen (stimmt nicht - stimmt kaum - stimmt eher - stimmt genau) bereit. Die Belastungs-Skala besteht aus den sechs Items der Subskala „Arbeitsüberforderung“ der Skala „Berufliche Belastung“ von Enzmann & Kleiber (1989) und kann einen Reliabilitätskoeffizienten von Cronbachs Alpha = .82 vorweisen. Die Skala „Schulklima“ wurde eigens für die Befragung erstellt und besteht aus insgesamt acht Items, die sich auf das Verhältnis zum Kollegium und zur Schulleitung beziehen und

darüber hinaus die schulinterne Fortbildung und das allgemeine „Wohlfühlen“ thematisieren. Der Reliabilitätskoeffizient für die Gesamtskala liegt bei Cronbachs Alpha = .76.

Der zweite Teil des Fragebogens beinhaltet zwei Fragen mit Antwortvorgaben sowie sieben Fragen, die eine Einschätzung von Items auf vierstufigen Skalen (stimmt nicht - stimmt kaum - stimmt eher - stimmt genau) vorsehen. Im Zentrum stehen nach der Frage, für wie bedeutsam, zutreffend und bedenklich die Ergebnisse der TIMS-Studie eingeschätzt werden und ob dieselben so erwartet wurden, die möglichen Ursachen des nur mittelmäßigen Abschneidens der deutschen Schülerinnen und Schüler bei TIMSS. Sie lassen sich zu den fünf Skalen „Ursache Eltern“ (z. B. „Es liegt daran, dass sich die Eltern zu wenig um die schulischen Leistungen kümmern“), „Ursache Gesellschaft“ (z. B. Ein Grund des schlechten Abschneidens liegt darin, dass in unserer Gesellschaft insgesamt nicht so viel geleistet und gearbeitet wird“), „Ursache Lehrerinnen und Lehrer“ (z. B. „Es liegt daran, dass die Lehrer/innen sich zu wenig anstrengen“), „Ursache Schülerinnen und Schüler“ (z. B. „Es liegt daran, dass die Schüler/innen im Unterricht nicht genügend aufpassen“) sowie „Ursache Schulverwaltung und Bildungspolitik“ (z. B. „Es liegt daran, dass zu oft Unterricht ausfällt“) gruppieren. Zusammen ergeben sich hier 60 Items, die in einer Voruntersuchung mit 10 Lehrkräften sowie 10 Eltern („Wie lässt sich Ihrer Meinung nach das nur mittelmäßige Abschneiden der deutschen Schülerinnen und Schüler bei TIMSS erklären?“) zusammengetragen worden waren. Die Reliabilitätskoeffizienten für die Einzelskalen liegen mit einer Ausnahme (Cronbachs Alpha = .61 für die aus nur vier Items bestehende Skala „Ursache Gesellschaft“) zwischen Cronbachs Alpha = .78 und .82. Anschließend wird erfragt, ob bzw. inwieweit die Lehrerinnen und Lehrer einen Bezug zu ihrer eigenen Situation herstellen (Skala „Mein Unterricht“ mit drei Items und Cronbachs Alpha = .57 sowie Items zu den vermuteten Leistungen Baden-Württembergs und auch der Grundschülerinnen und Grundschüler), ob sie an einer Veränderung und Verbesserung ihres Unterrichts im Hinblick auf die Ergebnisse von TIMSS interessiert sind (Skala „Veränderung“ mit Cronbachs Alpha = .77), ob bzw. unter welchen Umständen sie selbst bereit wären, an einer entsprechenden Studie mitzuwirken und wie sie generell zu externen Evaluationsstudien stehen (Skala „Untersuchungen“ mit fünf Items und Cronbachs Alpha = .81).

Um Informationen über die quantitativen Daten hinaus gewinnen zu können, beinhaltet die zweite Hälfte des Fragebogens auch fünf offen zu beantwortende Fragen (z. B. „Welche Gedanken gingen Ihnen durch den Kopf, als Sie von den Ergebnissen der TIMS-Studie hörten?“). Die Auswertung der hier erhobenen qualitativen Daten ist aber einem späteren Zeitpunkt vorbehalten.

3. Ergebnisse

Bekanntheit der Studie und Informiertheit der Lehrkräfte. Zunächst sollte mit Hilfe der hier berichteten Untersuchung festgestellt werden, inwieweit

die TIMS-Studie den befragten Lehrerinnen und Lehrern bekannt war bzw. für wie informiert sich die Lehrkräfte hielten. Es zeigte sich, dass sich von den Befragten 47 % bezüglich der TIMS-Studie für nicht informiert hielten, während sich nur 3 % als gut informiert bezeichneten. Selbst nach erfolgter Information gaben noch 15 % der Lehrkräfte an, noch nie etwas von TIMSS gehört zu haben. Da einige Schulleitungen bei der Ankündigung der Befragung nicht ganz auf eine Erläuterung, um was es sich bei TIMSS denn handelte, hatten verzichten wollen, kamen noch acht Personen hinzu, die mit der TIMS-Studie lediglich in Zusammenhang mit der hier beschriebenen Befragung in Kontakt getreten waren. Nimmt man diese Lehrkräfte hinzu, so erhöht sich der Anteil der vor der Befragung gänzlich uninformatierten Lehrerinnen und Lehrer auf knapp 20 %.

Die eindeutig wichtigste Informationsquelle der Informierten stellten Tageszeitung, Radio und Fernsehen mit 101 Nennungen (Mehrfachnennungen möglich) dar. Mit großem Abstand folgten informelle Gespräche an der Schule (53 Nennungen) und schließlich die offiziellen Informationsquellen für den Schuldienst in Form von Mitteilungen der (Ober-)Schulämter und des Ministeriums (41 Nennungen), in Form von Fachzeitschriften (35 Nennungen), Mitteilungen von Verbänden (29 Nennungen) sowie in Form von schulexterner (24 Nennungen) oder schulinterner (22 Nennungen) Fortbildung. Die am Max-Planck-Institut in Berlin entwickelte und auf Betreiben des baden-württembergischen Kultusministeriums im Jahr 1999 an alle Sekundarschulen des Landes einschließlich der Grund- und Hauptschulen sowie der Berufsschulen versandte CD-ROM „Mathematikunterricht der Sekundarstufe I in Deutschland, Japan und den USA - Dokumentation zur TIMSS-Videostudie“ (Klieme, Knoll & Schümer, 1999), die informieren und Diskussionen und Veränderungen anstoßen sollte, wird unter der Rubrik „zusätzliche Informationswege“ kein einziges Mal genannt. Auch bei der ausformulierten Frage nach einer etwaigen Kenntnis der CD-ROM im Anschluss an die Informationseinheit konnten nur 16 Befragte mit „Ja“ antworten, davon drei Schulleiterinnen bzw. Schulleiter, welche die unmittelbaren Adressaten der CD-ROMs waren.

Beim subjektiven Informationsstand, also bei der Frage, für wie informiert sich die Lehrerinnen und Lehrer selbst hielten, zeigten sich klare Unterschiede zwischen den Lehrerinnen und Lehrern mit verschiedener Fächerzugehörigkeit. Erwartungsgemäß fiel der subjektive Informationsstand bei jenen Lehrkräften am höchsten aus, die nur Mathematik und/oder Naturwissenschaften unterrichteten und bei jenen am niedrigsten, die weder Mathematik noch Naturwissenschaften unterrichteten ($F(2,196) = 16.80, p < .001$); vgl. Tabelle 1). Gleiches galt für das tatsächlich erhobene Vorwissen, das bei den entsprechenden Fachlehrerinnen und Fachlehrern am höchsten und bei den Lehrkräften, die weder Mathematik noch Naturwissenschaften unterrichteten, am geringsten ausfiel ($F(2,197) = 13.40, p < .001$; vgl. Tabelle 1). Auffällig bei den Lehrerinnen und Lehrern aller Schularten waren hier die großen Streuungen. Eine zusätzliche Unterscheidung nach der studierten Fächer-

Tabelle 1:
Mittelwerte (Standardabweichungen) von subjektivem Informationsstand und erhobenem Vorwissen in Abhängigkeit von Fächerkombination und Schulart

Schulart	Inform.stand	<i>n</i>	Vorwissen	<i>n</i>
Unterrichtete Fächer: Nur Mathematik und/oder Naturwissenschaften				
Grund- und Hauptschule	2.00 (.82)	4	3.50 (2.38)	4
Realschule	1.86 (1.10)	7	3.57 (3.05)	7
Gymnasium	2.23 (.83)	44	5.48 (2.53)	44
Gesamt	2.16 (.86)	55	5.09 (2.66)	55
Unterrichtete Fächer: Nicht Mathematik und/oder Naturwissenschaften				
Grund- und Hauptschule	1.00 (.00)	9	0.00 (.00)	9
Realschule	1.32 (.58)	19	2.65 (2.94)	20
Gymnasium	1.46 (.58)	26	3.42 (2.44)	26
Gesamt	1.33 (.55)	54	2.58 (2.69)	55
Unterrichtete Fächer: Auch Mathematik und/oder Naturwissenschaften				
Grund- und Hauptschule	1.31 (.47)	48	1.90 (2.62)	48
Realschule	2.19 (.98)	26	4.77 (2.60)	26
Gymnasium	1.94 (.68)	16	3.75 (2.35)	16
Gesamt	1.68 (.79)	90	3.06 (2.85)	90

kombination erwies sich sowohl bezüglich des subjektiven Informationsstandes als auch bezüglich des Vorwissens als überflüssig, da sie praktisch identische Resultate wie die Differenzierung nach den unterrichteten Fächern erbrachte.

Subjektiver Informationsstand und tatsächlich erhobenes Vorwissen unterschieden sich auch in Abhängigkeit von der unterrichteten Schulart deutlich. Wie Tabelle 2 entnommen werden kann, fiel der subjektive Informationsstand bei den Lehrerinnen und Lehrern an Grund- und Hauptschulen am geringsten und bei den Lehrkräften an Gymnasien am höchsten aus ($F(2,196) = 12.76, p < .001$). Die Realschullehrerinnen und -lehrer standen den Gymnasiallehrkräften jedoch nur in unbedeutender Weise nach, während sie sich von den Grund- und Hauptschullehrerinnen und -lehrern signifikant unter-

Tabelle 2:
Deskriptive Kennwerte für den subjektiven Informationsstand
in Abhängigkeit von der Schulart

Schulart	<u>M</u>	<u>SD</u>	Min	Max
Grund- und Hauptschule (<i>n</i> = 61)	1.31	.50	1	3
Realschule (<i>n</i> = 52)	1.83	.94	1	4
Gymnasium (<i>n</i> = 86)	1.94	.80	1	4
Gesamt (<i>n</i> = 199)	1.72	.81	1	4

schieden ($p < .01$). Von den letzteren hielten sich 70 % für nicht informiert. Vergleichbares ergab sich im Hinblick auf das erhobene Vorwissen, das wiederum bei den Grund- und Hauptschullehrkräften am geringsten und bei den Lehrerinnen und Lehrern an Gymnasien am höchsten war ($F(2,196) = 19.60, p < .001$; vgl. Tabelle 3). 62 % der Grund- und Hauptschullehrkräfte erzielten 0 Punkte bei einer maximal erreichbaren Anzahl von 8 Punkten. Sie

Tabelle 3:
Deskriptive Kennwerte für das erhobene Vorwissen
in Abhängigkeit von der Schulart

Schulart	<u>M</u>	<u>SD</u>	Min	Max
Grund- und Hauptschule (<i>n</i> = 61)	1.72	2.52	0	8
Realschule (<i>n</i> = 53)	3.81	2.91	0	8
Gymnasium (<i>n</i> = 86)	4.54	2.63	0	8
Gesamt (<i>n</i> = 200)	3.49	2.92	0	8

unterschieden sich signifikant von den Realschullehrkräften ($p < .001$), deren Differenz zu den Gymnasiallehrkräften sich bei einem Einzelvergleich als statistisch nicht bedeutsam erwies.

Weil die Ziehung einer ausbalancierten Stichprobe nicht möglich war, konnte eine Konfundierung der Effekte von Schulart und Fächerkombination nicht vermieden werden. Auffällig war aber, dass die Lehrerinnen und Lehrer an Grund- und Hauptschulen auch bei einer nach der unterrichteten Fächerkombination unterschiedenen Betrachtung nahezu immer die schwächsten Ergebnisse erzielten (vgl. Tabelle 1). Dies bedeutete beispielsweise, dass selbst Gymnasiallehrerinnen und -lehrer, die weder Mathematik noch Naturwissenschaften unterrichteten, den entsprechenden Kolleginnen und Kollegen an Grund- und Hauptschulen in ihrem Vorwissen überlegen waren ($p < .01$). Bei einer zweifaktoriellen Varianzanalyse erwiesen sich die Faktoren Schulart ($F(2,192) = 7.77, p < .01$) und Fächerkombination ($F(2,192) = 7.12, p < .01$) im Hinblick auf das tatsächlich vorhandene Vorwissen als etwa gleich relevant, während sich hinsichtlich des subjektiven Informationsstandes die Frage der Fächerkombination ($F(2,190) = 11.90, p < .001$) als gewichtiger als die jeweilige Schulart ($F(2,190) = 3.45, p < .05$) zeigte. Aufgrund einer sehr unterschiedlichen und in einem Fall auch sehr kleinen Zellenbesetzung ($n = 4$) sind diese Ergebnisse jedoch mit Vorsicht zu betrachten.

Die Frage nach den vorhandenen Zusatzqualifikationen und/oder ausgeübten Zusatzfunktionen erwies sich bei der Gesamtstichprobe sowohl im Hinblick auf den subjektiven Informationsstand ($r = .12, n. s.$) als auch bezüglich des tatsächlich vorhandenen Vorwissens ($r = .12, n. s.$) als bedeutungslos. Ein deutlicher Zusammenhang zwischen subjektivem Informationsstand und tatsächlich erhobenem Vorwissen konnte dagegen ermittelt werden ($r = .62, p < .001$).

Einschätzung der Ergebnisse der TIMS-Studie. Bei der Frage, für wie bedeutsam ($M = 2.66, SD = .89$) und zutreffend ($M = 2.78, SD = .78$) die Ergebnisse der TIMS-Studie gehalten wurden, ergaben sich Ratings im durchschnittlichen Bereich der vierstufigen Skala. Bezüglich der Einschätzung der Bedeutsamkeit zeigten sich keine signifikanten Unterschiede in Abhängigkeit von der Schulart oder der studierten bzw. unterrichteten Fächerkombination. Auch waren bedeutsame Zusammenhänge mit dem subjektiven Informationsstand, dem erhobenen Vorwissen sowie der Frage, ob zusätzliche Qualifikationen vorlagen oder besondere Funktionen ausgeübt wurden, nicht feststellbar.

Im Hinblick auf die Frage, für wie zutreffend die Ergebnisse gehalten wurden, zeigte eine einfaktorielle Varianzanalyse jedoch, dass sich die Lehrkräfte der verschiedenen Schularten bedeutsam voneinander unterschieden ($F(2,185) = 3.18, p < .05$). Dies galt in der Weise, dass die Grund- und Hauptschullehrkräfte ($M = 3.00, SD = .71$) die Ergebnisse im Vergleich zu ihren Kolleginnen und Kollegen an Realschulen ($M = 2.68, SD = .90$) und Gymnasien ($M = 2.69, SD = .73$) als zutreffender einstufen, wobei die Einzelvergleiche die Signifikanzgrenze jeweils knapp verfehlten. Ein Zusammenhang

zwischen der Einschätzung der Ergebnisse nach ihrer Bedeutsamkeit und der Frage, für wie zutreffend diese gehalten wurden, war deutlich erkennbar ($r = .58, p < .001$).

Das Item, das eruieren sollte, ob das Abschneiden der deutschen Schülerinnen und Schüler bei TIMSS für bedenklich gehalten wurde, erhielt von der Gesamtstichprobe eine Einschätzung unweit des theoretischen Mittelwertes ($M = 2.88, SD = .88$). Das Abschneiden wurde als umso bedenklicher eingestuft, je bedeutsamer ($r = .40, p < .001$) bzw. zutreffender ($r = .32, p < .001$) die Ergebnisse der Studie eingeschätzt wurden.

Bei der Frage, ob das Abschneiden in der festgestellten Weise erwartet worden war, zeigten sich die Befragten unentschieden ($M = 2.53, SD = .85$), wobei dies für die Lehrerinnen und Lehrer aller Schularten in etwa gleichem Maße galt.

Haltung zu (kultur-)vergleichenden Leistungsstudien. Die Skepsis vergleichenden Leistungsstudien gegenüber fiel eher gering aus ($M = 2.06, SD = .61$). Auch dem Einzelitem „Die Untersuchungsergebnisse sind wertlos, weil die Wissenschaftler an der Hochschule keine Ahnung haben, was in der Schule passiert“ wurde wenig zugestimmt ($M = 2.04, SD = .84$). Die größte Schwierigkeit in diesem Zusammenhang schienen die Lehrerinnen und Lehrer mit vergleichenden Leistungsstudien internationaler Prägung zu haben: Das Item „Zu den Ergebnissen ist zu sagen, dass man die Leistungen von Schüler/innen aus verschiedenen Ländern gar nicht sinnvoll miteinander vergleichen kann“ erfuhr von den hier relevanten fünf Items die eindeutig höchste Zustimmung ($M = 2.50, SD = .83$).

Es ergaben sich keine bedeutsamen Unterschiede in Abhängigkeit von der Schulart. Auch das Alter der Lehrkräfte und die Frage, ob sie über Zusatzqualifikationen bzw. Zusatzfunktionen verfügten, spielten hier keine Rolle. Zusammenhänge zeigten sich hingegen mit den Einschätzungen der Ergebnisse der TIMS-Studie: Je kritischer die Lehrerinnen und Lehrer externer Evaluation gegenüberstanden, desto weniger hielten sie auch die Ergebnisse der TIMS-Studie für bedeutsam ($r = -.46, p < .001$), zutreffend ($r = -.42, p < .001$) und bedenklich ($r = -.33, p < .001$). Signifikante, aber schwache Korrelationen ergaben sich darüber hinaus mit den Skalen „Ursache Eltern“ ($r = .22, p < .01$), „Ursache Gesellschaft“ ($r = .25, p < .01$), „Ursache Schülerinnen und Schüler“ ($r = .27, p < .001$) und „Ursache Schulverwaltung und Bildungspolitik“ ($r = .32, p < .001$) - nicht aber mit der Skala „Ursache Lehrerinnen und Lehrer“ ($r = .01, n. s.$). Dies bedeutet, dass eine kritischere Einstellung gegenüber externen Evaluationsstudien mit der Tendenz einherging, das wenig erfreuliche Abschneiden der deutschen Schülerinnen und Schüler bei TIMSS stärker mit Ursachen außerhalb der eigenen Berufsgruppe zu begründen, also eine verstärkte externe Ursachenzuschreibung vorzunehmen.

Frage nach den möglichen Ursachen des nur mittelmäßigen Abschneidens der deutschen Schülerinnen und Schüler. Bei der Frage nach den möglichen

Ursachen des nur mittelmäßigen Abschneidens der deutschen Schülerinnen und Schüler bietet der eingesetzte Fragebogen 60 Items an, die sich zu den fünf Skalen „Ursache Eltern“, „Ursache Gesellschaft“, „Ursache Lehrerinnen und Lehrer“, „Ursache Schülerinnen und Schüler“ sowie „Ursache Schulverwaltung und Bildungspolitik“ gruppieren lassen. Die höchste Zustimmung unter den Befragten erzielte die Skala „Ursache Schülerinnen und Schüler“, während die geringste Zustimmung der Skala „Ursache Lehrerinnen und Lehrer“ zuteil wurde. Auch wenn man bedenkt, dass die Skala „Ursache Schulverwaltung und Bildungspolitik“ eine ähnlich verhaltene Zustimmung wie die Skala „Ursache Lehrerinnen und Lehrer“ erfuhr (vgl. Tabelle 4), so ergibt sich hier dennoch ein Hinweis auf typische Attributionsvoreingenommenheiten, die sich in einer externalen Ursachenzuschreibung

Tabelle 4:
Deskriptive Kennwerte der Ratings zu den „Ursache“-Skalen

Skala	<u>M</u>	<u>SD</u>	<u>n</u>
Ursache Eltern	2.70	.50	187
Ursache Gesellschaft	2.90	.50	197
Ursache Lehrerinnen und Lehrer	2.29	.38	187
Ursache Schülerinnen und Schüler	3.07	.43	192
Ursache Schulverwaltung und Bildungspolitik	2.31	.34	176

bei der Erklärung von Misserfolg äußern. Dieser Hinweis erfährt Bestärkung bei einem Vergleich der Ratings von Lehrkräften, die ausschließlich Mathematik und/oder Naturwissenschaften unterrichteten bzw. auch oder nicht unterrichteten. Wie Tabelle 5 entnommen werden kann, fiel die Zustimmung zur Skala „Ursache Lehrerinnen und Lehrer“ bei jenen Lehrkräften am geringsten aus, die ausschließlich Mathematik und/oder Naturwissenschaften unterrichteten, und bei jenen am höchsten, die weder in Mathematik noch in den Naturwissenschaften Unterricht erteilten. Eine einfaktorielle Varianzanalyse konnte zeigen, dass sich die drei Gruppen hier tatsächlich bedeutsam voneinander unterschieden ($F(2,184) = 5.82, p < .01$). Zusätzlich durchgeführte Einzelvergleiche erbrachten, dass sich die Mathematiker und Naturwissenschaftler signifikant von den Nichtmathematikern und Nichtnaturwissenschaftlern unterschieden ($p < .01$), während die beiden anderen Differenzen das Signifikanzniveau verfehlten. Bei allen anderen „Ursache“-Skalen ließ sich hingegen kein Einfluss der unterrichteten Fächer feststellen.

Tabelle 5:

Deskriptive Kennwerte der Ratings zur Skala „Ursache Lehrerinnen und Lehrer“ in Abhängigkeit von der unterrichteten Fächerkombination

Skala	<i>M</i>	<i>SD</i>	<i>n</i>
nur Mathematik und/oder Naturwissenschaften	2.19	.28	51
weder Mathematik noch Naturwissenschaften	2.43	.41	49
auch Mathematik und/oder Naturwissenschaften	2.27	.39	87

Eine besonders geringe Zustimmung erfuhren einzelne Items der Skalen „Ursache Lehrerinnen und Lehrer“ sowie „Ursache Schulverwaltung und Bildungspolitik“. Die höchsten Ratings erzielten dagegen Items der beiden Skalen „Ursache Gesellschaft“ und „Ursache Schülerinnen und Schüler“ (vgl. Tabelle 6). Die Items der Skala „Ursache Lehrerinnen und Lehrer“, die sich auf die Methodenkompetenz der Lehrkräfte und ihr Bemühen, Kreativität und Problemlösefähigkeit bei den Lernenden zu fördern, bezogen, erhielten Ratings, die nicht allzu weit um den theoretischen Mittelwert der Skala streuten (Item „Die schlechten Ergebnisse kommen dadurch zustande, dass die Lehrer/innen oftmals mit ungeeigneten Methoden unterrichten“ mit $M = 2.32$ und $SD = .72$; Item „Es liegt daran, dass im Unterricht zu wenig Wert auf Problemlösen gelegt wird“ mit $M = 2.90$ und $SD = .86$; Item „Es liegt daran, dass im Unterricht die Kreativität zu wenig gefördert wird“ mit $M = 2.86$ und $SD = .74$).

Die oben erwähnte vergleichsweise geringe Zustimmung zur Skala „Ursache Schulverwaltung und Bildungspolitik“ sollte nicht darüber hinwegtäuschen, dass einzelnen Items dieser Skala weniger verhalten zugestimmt wurde. Dies betrifft insbesondere das Item zur Klassengröße (vgl. Tabelle 6), aber auch noch das Item „Das schlechte Abschneiden der deutschen Schüler/innen ist das Ergebnis schlechter Bildungspolitik“ ($M = 2.94$, $SD = .70$) und, wenn auch in geringerem Maße, das Item „Eine Ursache besteht darin, dass die Lehrer/innen mit ihren Schwierigkeiten keine Unterstützung 'von oben' erfahren“ ($M = 2.74$, $SD = .85$). Die Kritik an der eigenen Ausbildung fiel gering aus: Das Item „Es liegt daran, dass die Lehrer/innen unzureichend ausgebildet sind“ ($M = 2.08$, $SD = .80$) erfuhr eine vergleichsweise geringe Zustimmung.

Tabelle 6:

Deskriptive Kennwerte der Items der „Ursache“-Skalen mit besonders geringer und besonders hoher Zustimmung

Item	<i>M</i>	<i>SD</i>	<i>n</i>
besonders geringe Zustimmung			
"Es liegt daran, dass sich die Lehrer/innen zu wenig anstrengen"	1.59	.62	201
"Es liegt daran, dass die Lehrer/innen die Schule selbst nicht wichtig genug nehmen"	1.70	.69	200
"Es liegt daran, dass die Schulbücher schlecht sind"	1.78	.67	197
"Ein Grund des schlechten Abschneidens liegt darin, dass viele Lehrer/innen fachfremd Mathematik und die Naturwissenschaften unterrichten und damit in diesen Fächern nicht so erfolgreich wirken"	1.83	.79	196
"Es liegt daran, dass die Kinder im Kindergarten zu wenig gefordert und damit auch zu wenig gefördert werden"	1.86	.78	194
"Es liegt an der gängigen Hausaufgabenpraxis, die keine Rücksicht auf benachteiligte Schüler/innen nimmt"	1.89	.70	198
"Es liegt daran, dass zu oft Unterricht ausfällt"	1.93	.77	201
"Eine Ursache liegt darin, dass der Lehrplan zu wenig verlangt"	1.95	.75	199
"Eine Ursache des schlechten Abschneidens liegt darin, dass die Lehrer/innen nicht bereit sind, sich auf die Probleme der heutigen Schüler/innen einzustellen"	1.97	.75	200
"Es liegt daran, dass es bei uns kaum Ganztagschulen gibt"	1.99	.93	197
besonders hohe Zustimmung			
"Es liegt daran, dass man in unserer Konsumgesellschaft nur noch Angebote wahrzunehmen braucht und nichts mehr selbst gestalten oder durchdenken muss"	3.45	.60	200
"Es liegt daran, dass die Schüler/innen heute zu wenig lesen und oft Schwierigkeiten haben, einem Text den Sinn zu entnehmen"	3.36	.63	201
"Ein Grund für die schlechten Ergebnisse liegt darin, dass die heutigen Schüler/innen sich kaum noch konzentrieren können"	3.25	.63	201
"Es liegt daran, dass die Leistungsbereitschaft der Schüler/innen zu gering ist"	3.16	.68	199
"Es liegt daran, dass sich die Schüler/innen in unserer Freizeitgesellschaft keine Zeit mehr zum häuslichen Lernen nehmen"	3.16	.65	200
"Es liegt daran, dass in unserem reichen und technisierten Staat die Kinder im Spiel kaum noch kreativ sind"	3.11	.71	200
"Ein Grund besteht darin, dass die Klassen zu groß sind"	3.09	.92	199

Die Ratings der Befragten unterschieden sich zum Teil in Abhängigkeit von der jeweiligen Schulart: Die Gymnasiallehrerinnen und -lehrer stimmten den Items der Skala „Ursache Schülerinnen und Schüler“ ($M = 2.98$, $SD = .46$; $F(2,184) = 9.51$, $p < .001$) im Vergleich zu ihren Kolleginnen und Kollegen an den beiden anderen Schularten ($M = 3.17$, $SD = .45$ für die Realschullehrkräfte; $M = 3.13$, $SD = .36$ für die Grund- und Hauptschullehrkräfte) etwas weniger stark zu. Gleiches gilt für die Items der Skala „Ursache Eltern“ ($F(2,189) = 3.57$, $p < .05$), denen vonseiten der Gymnasiallehrkräfte ($M = 2.53$, $SD = .49$) weniger Zustimmung zuteil wurde als vonseiten der Lehrkräfte an Realschulen ($M = 2.80$, $SD = .51$) und an Grund- und Hauptschulen ($M = 2.87$, $SD = .44$). Der Einfluss des Alters auf die Ratings zeigte sich dagegen bei den Skalen „Ursache Gesellschaft“ und „Ursache Lehrerinnen und Lehrer“: Mit zunehmendem Alter standen Lehrerinnen und Lehrer im Kontext dieser Befragung der heutigen Gesellschaft offenbar etwas kritischer gegenüber ($r = .22$, $p < .01$), und jüngere Lehrkräfte unterstellten ihren Kolleginnen und Kollegen einen tendenziell höheren Anteil am schwachen Testergebnis ($r = -.22$, $p < .01$). Bei der Frage nach den möglichen Ursachen der TIMSS-Testergebnisse kam der im Beruf empfundenen Belastung eine eher schwache Bedeutung zu: Zusammenhänge mit der empfundenen Belastung zeigten sich in sehr geringem bis geringem Maße bei den Skalen „Ursache Gesellschaft“ ($r = .16$, $p < .05$), „Ursache Schülerinnen und Schüler“ ($r = .24$, $p < .01$), „Ursache Lehrerinnen und Lehrer“ ($r = .19$, $p < .05$) sowie „Ursache Schulverwaltung und Bildungspolitik“ ($r = .33$, $p < .001$), wobei eine höhere Belastung zu einem höheren Rating führte.

Nicht bedeutsam in diesem Zusammenhang war die Frage, ob die Personen über Zusatzqualifikationen verfügten oder zusätzliche Funktionen ausübten. Gleiches galt fast durchgängig für den subjektiven Informationsstand und das vorhandene Vorwissen. Beide Variablen korrelierten lediglich mit der Skala „Ursache Eltern“ in geringem Maße ($r = -.18$, $p < .05$ für den subjektiven Informationsstand; $r = -.14$, $p < .05$ für das Vorwissen). Schließlich spielte es im Hinblick auf die hier interessierenden Ratings keine Rolle, ob die Befragten selbst Kinder hatten, also sowohl die Lehrer- als auch die Elternrolle einnahmen.

Bezug zur eigenen Situation. Da die befragten Lehrerinnen und Lehrer - wie alle ihre Kolleginnen und Kollegen in Baden-Württemberg - nicht an TIMSS teilgenommen hatten, erschien es interessant zu erfahren, ob bzw. inwieweit von den Befragten ein Bezug von den Ergebnissen der TIMS-Studie zur eigenen Situation hergestellt werden würde. Es zeigte sich, dass die Lehrkräfte eher weniger glaubten - ohne diese Möglichkeit völlig von der Hand zu weisen -, dass Baden-Württemberg besser als die übrigen Bundesländer abgeschnitten hätte ($M = 2.13$, $SD = .69$). Bei dieser Frage war eine schulartspezifische Einschätzung dergestalt zu finden, dass Gymnasiallehrkräfte ($M = 2.31$, $SD = .73$) ihrem Bundesland am meisten zutrauten, während die Grund- und Hauptschullehrkräfte ($M = 1.90$, $SD = .64$) hier am skeptischsten erschienen ($F(2,18) = 6.51$, $p < .01$). Daneben zeigte sich, dass Lehrkräfte die

Leistungen ihres eigenen Bundeslandes mit zunehmendem Alter etwas höher einschätzten ($r = .25, p < .001$). Signifikante Zusammenhänge ergaben sich weiterhin mit den Skalen „Untersuchung“ ($r = .40, p < .001$) und „Ursache Gesellschaft“ ($r = .24, p < .01$). Dies bedeutet, dass eine höhere Meinung vom eigenen Bundesland einherging mit einer höheren Skepsis gegenüber externer Evaluation und mit einer kritischeren Einstellung gegenüber potentiell leistungsrelevanten gesellschaftlichen Zuständen.

Einen ähnlichen Wert wie die Items zum vermuteten Abschneiden Baden-Württembergs erzielte das Item „Hätte man die Leistungen der Grundschüler/innen untersucht, wären die Ergebnisse nicht so schlecht ausgefallen“ ($M = 2.12, SD = .88$). Auch hier zeigten sich bei einer einfaktoriellen Varianzanalyse wieder schulartspezifische Unterschiede, wenngleich dieses Mal unter anderem Vorzeichen ($F(2, 186) = 7.89, p < .001$): Nun waren es die Grund- und Hauptschullehrerinnen und -lehrer ($M = 2.46, SD = .89$), die diesem Item deutlich mehr zustimmten als ihre Kolleginnen und Kollegen an Realschulen ($M = 2.06, SD = .84$) und Gymnasien ($M = 1.89, SD = .83$).

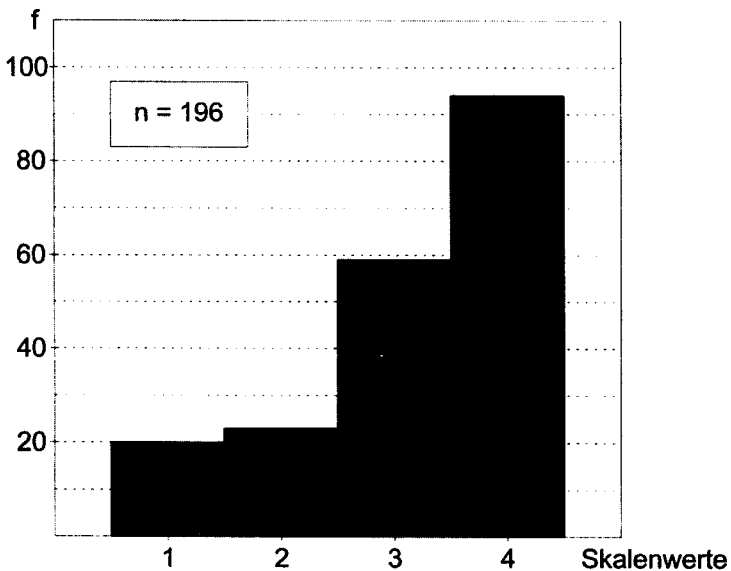
Bei der Frage, ob den Ergebnissen der TIMS-Studie auch eine - unerfreuliche - Aussage über den eigenen Unterricht entnommen werden kann, ergab sich eine Einschätzung nahe am theoretischen Mittelwert ($M = 2.39, SD = .68$). Ein schwacher, aber signifikanter Zusammenhang ergab sich dabei mit der Skala „Untersuchung“ in der Weise, dass eine höhere Ablehnung externer Evaluationsstudien einherging mit einer besseren Meinung vom eigenen Unterricht ($r = .25, p < .01$). Je positiver die Befragten dagegen das Klima ihrer Schule einschätzten ($r = -.20, p < .01$) und je mehr Vorwissen über die TIMS-Studie sie mitbrachten ($r = .17, p < .05$), desto kritischer standen sie ihrem eigenen Unterricht gegenüber. Ein bedeutsamer Zusammenhang mit dem Alter der Befragten ($r = .07, n. s.$) konnte dagegen nicht ermittelt werden.

Bereitschaft zur Veränderung. Die in den Fragebögen signalisierte Bereitschaft zur Veränderung und Verbesserung des eigenen Unterrichts fiel relativ hoch aus (M für die Skala „Veränderung“ = $2.98, SD = .55$). Besondere Zustimmung bei den Befragten erfuhr dabei das Item „Ich habe Interesse an einer Fortbildung, die zeigt, wie ich meinen Unterricht im Hinblick auf TIMSS positiv verändern kann“ ($M = 3.15, SD = .91$) und noch mehr das Item „Ich habe Interesse an neuem Unterrichtsmaterial, mit dem ich meinen Unterricht im Hinblick auf TIMSS ohne viel Aufwand verbessern kann“ ($M = 3.30, SD = .83$). Zu beachten ist hier allerdings, dass lediglich die Bereitschaft zu einer Veränderung erfragt wurde - und nicht tatsächlich durchgeführte Maßnahmen. Die Bereitschaft zur Veränderung und Verbesserung des eigenen Unterrichts wuchs in dem Maße, in dem die Ergebnisse der TIMS-Studie für bedeutsam ($r = .36, p < .001$), zutreffend ($r = .34, p < .001$) und bedenklich ($r = .24, p < .01$) gehalten wurden, und in geringerer Weise auch in dem Maße, in dem sich die Befragten wohl an ihrer Schule fühlten ($r = .19, p < .05$). Sie verringerte sich mit höherer Skepsis externen Evaluationsstudien gegen-

über ($r = -.56, p < .001$) und auch dann, wenn die Ursachen des wenig erfreulichen Abschneidens der deutschen Schülerinnen und Schüler mehr bei den Lernenden ($r = -.19, p < .05$) und deren Eltern ($r = -.22, p < .01$), bei der Schulverwaltung und der Bildungspolitik ($r = -.19, p < .05$) sowie der Gesellschaft ($r = -.19, p < .05$) gesehen wurden, somit also eine stärkere externe Ursachenzuschreibung vorlag. Ein Zusammenhang mit der Einschätzung von Items der Skala „Ursache Lehrerinnen und Lehrer“ konnte dagegen nicht ermittelt werden ($r = .03, n. s.$). Schwache negative Korrelationen ergaben sich weiterhin mit dem Alter der Befragten ($r = -.19, p < .05$) sowie der empfundenen Belastung ($r = -.22, p < .01$), die mit dem Alter praktisch nicht korrelierte ($r = .01, n. s.$). Mit höher empfundener Belastung stimmten die Lehrkräfte insbesondere den Items „Bevor ich aktiv werde, müssen erst einmal Kultusministerium und Schulverwaltung etwas tun“ ($r = -.27$ (umgepolt), $p < .001$) und „Auch wenn ich mich sehr bemühe, kann ich meine Schüler/innen doch nicht zu höheren Denkleistungen führen“ ($r = -.27$ (umgepolt), $p < .001$) zu.

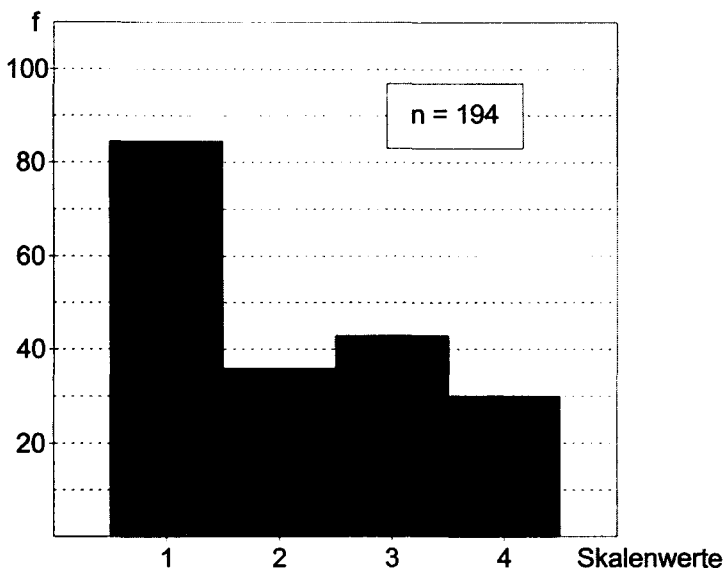
Bereitschaft zur Teilnahme an externen Evaluationsstudien. Die Bereitschaft zur Teilnahme an externen Evaluationsstudien fiel unter der Bedingung von Anonymität hoch aus ($M = 3.12, SD = .99$; vgl. Abbildung 1). Kön-

Abbildung 1:
Zustimmung zum Item „An einer solchen Studie würde ich bei absoluter Anonymität teilnehmen“



te bei einer Studie Anonymität hingegen nicht zugesichert werden, so würde die Teilnahmebereitschaft deutlich absinken ($M = 2.09, SD = 1.13$; vgl. Abbildung 2). Die Teilnahmebereitschaft unter der Bedingung von Anonymität hing zusammen mit der generellen Haltung vergleichenden Leistungsstudien

Abbildung 2:
Zustimmung zum Item „An einer solchen Studie würde ich auch ohne Zusicherung
von Anonymität teilnehmen“



gegenüber ($r = -.28, p < .001$). Die höchste Korrelation ergab sich dabei mit dem Item „Solchen Untersuchungen kann man sowieso nicht trauen“ ($r = -.40, p < .001$), die geringste mit dem Item „Das schlechte Ergebnis kam dadurch zustande, dass die Testaufgaben ungünstig gewählt waren“ ($r = -.11, n. s.$). Personen, die - unter der Bedingung von Anonymität - eher bereit zur Teilnahme an leistungsvergleichenden Studien waren, zeigten auch ein höheres Interesse, ihren Unterricht zu verändern und zu verbessern ($r = .33, p < .001$). Sie wären auch ohne Zusicherung von Anonymität eher bereit, an einer leistungsvergleichenden Studie teilzunehmen ($r = .46, p < .001$).

Das Interesse bezüglich des Abschneidens der eigenen Schülerinnen und Schüler bei einer vergleichenden Leistungsstudie fiel hoch aus ($M = 3.38, SD = .92$). Dabei ergab sich auch hier ein signifikanter Zusammenhang mit der allgemeinen Haltung externen Evaluationsstudien gegenüber ($r = -.33, p < .001$): Je höher die Skepsis leistungsvergleichenden Studien gegenüber ausfiel, desto geringer erschien das Interesse an entsprechenden Ergebnissen. Die höchste Korrelation bestand wiederum mit dem Item „Solchen Untersuchungen kann man sowieso nicht trauen“ ($r = -.43, p < .001$), und die geringste mit dem Item „Das schlechte Abschneiden kam dadurch zustande, dass die Testaufgaben ungünstig gewählt waren“ ($r = -.13, n. s.$). Das Interesse bezüglich des Abschneidens der eigenen Schülerinnen und Schüler korrelierte deutlich mit der jeweiligen Bereitschaft, an Studien externer Evaluation teilzunehmen ($r = .74, p < .001$). Schwache Zusammenhänge ergaben sich darüber hinaus mit den Skalen „Baden-Württemberg“ ($r = -.24, p < .01$).

sowie „Mein Unterricht“ ($r = -.32, p < .001$). Dies bedeutet, dass ein höheres Interesse an möglichen Untersuchungsergebnissen einherging mit einer kritischeren Haltung gegenüber den Leistungen baden-württembergischer Schülerinnen und Schüler im Allgemeinen und der eigenen Schülerinnen und Schüler im Besonderen.

Ein Extremgruppenvergleich, der Personen mit hoher vs. geringer Skepsis gegenüber externer Evaluation einander gegenüberstellte, kann diese sowie weiter oben berichtete Befunde noch einmal verdeutlichen (Tabelle 7). Interessant in diesem Zusammenhang erscheint, dass sich die beiden Extremgruppen mit Hilfe der in der Untersuchung erhobenen personenbezogenen Daten

Tabelle 7:
Extremgruppenvergleich. Mittelwerte (Standardabweichungen) für Personen
mit hoher vs. geringer Skepsis gegenüber externer Evaluation

Variable	Vpn hoch skept. <i>n</i>		Vpn gering skept. <i>n</i>	
Bereitschaft Teilnahme anonym*	2.82 (1.08)	44	3.54 (.75)	52
Interesse an Ergebnissen**	3.00 (1.08)	44	3.73 (.64)	51
Einschätzung TIMSS bedeutsam**	2.09 (.78)	43	3.16 (.82)	50
Einschätzung TIMSS zutreffend**	2.32 (.69)	41	3.15 (.72)	47
Abschneiden TIMSS bedenklich*	2.58 (.96)	45	3.22 (.79)	50
Skala "Ursache Schulverw. u. B."*	2.38 (.36)	43	2.16 (.33)	45
Skala "Ursache Eltern"*	2.86 (.49)	44	2.61 (.56)	49
Skala "Ursache Gesellschaft"*	3.13 (.41)	44	2.75 (.59)	52
Skala "Ursache Lehrerinnen u. L."	2.26 (.44)	44	2.34 (.37)	46
Skala "Ursache Schülerinn. u. S."*	3.24 (.35)	45	3.00 (.52)	49
Skala "Baden-Württemberg"***	2.50 (.74)	45	1.80 (.61)	50
Skala "Mein Unterricht"*	2.63 (.61)	45	2.25 (.76)	52
Skala "Veränderung"***	2.34 (.61)	35	3.37 (.70)	44
Vorwissen	3.62 (2.72)	45	3.52 (3.17)	52

*: $p < .01$ **: $p < .001$

Anmerkung. Die Datenlage ließ für den Extremgruppenvergleich keine Aufteilung der Gesamtstichprobe ($n = 178$) in drei gleich große Gruppen zu.

nicht unterscheiden ließen, sie also bezüglich Alter, Schulart, Belastung, Zusatzqualifikationen bzw. Zusatzfunktionen sowie Fächerkombination vergleichbar erschienen. Der einzige signifikante Unterschied ($F(2,163) = 3.48, p < .05$) zwischen den beiden Gruppen betraf die Integration in Schule und Kollegium, wobei die Befragten mit geringer Skepsis ($M = 2.84$,

$SD = .56$) sich an ihrer Schule und in ihrem Kollegium etwas wohler fühlten als die Lehrkräfte mit hoher Skepsis ($M = 3.11$, $SD = .55$). Ein Unterschied im erhobenen Vorwissen war zwischen den beiden Extremgruppen nicht feststellbar.

4. Diskussion

Zunächst einmal fällt auf, dass etwa ein Fünftel der Lehrerinnen und Lehrer angab, vor der hier beschriebenen Befragung noch nie etwas von der TIMS-Studie gehört zu haben. Darüber hinaus zeigten auch jene Lehrerinnen und Lehrer, denen die TIMS-Studie bereits bekannt war, im Mittel einen sehr niedrigen Informationsstand. Dieses Ergebnis kann schon als bemerkenswert bezeichnet werden - vor allem wenn man bedenkt, wie verbreitet und wie heftig die Diskussionen über die TIMS-Studie sind und waren und welche große Zahl an Informationskanälen genutzt werden konnte.

Betrachtet man sich die von den Informierten genutzten Informationsquellen, so zeigt sich, dass die offiziellen Organe für den Schuldienst erst an dritter Stelle genannt wurden bzw. sie vier Fünftel der Lehrerschaft nicht erreichten. Auch wenn man in Rechnung stellt, dass sich viele Befragte wohl einfach nicht mehr daran erinnern konnten, vom Ministerium oder Schulamt entsprechende Informationen erhalten zu haben, so wird doch deutlich, wie ineffektiv dieser Weg der Informationsweitergabe zumindest in diesem Fall war. Noch mehr gilt dies für eine mit einem Anschreiben und einer Broschüre im Jahr 1999 an alle Sekundarschulen des Landes verschickte CD-ROM, die explizit den Auftrag hatte, für differenziertes Wissen über die TIMS-Studie zu sorgen und entsprechende Diskussionen anzustoßen: Sie war praktisch gar nicht wahrgenommen worden. Leider lässt sich aufgrund der vorliegenden quantitativen Daten und auch im Nachhinein nicht entscheiden, warum die Informationen bei den Adressaten nicht angekommen waren, ob sie also den Weg von den Schulleitungen zu den Kollegien nicht gefunden hatten oder ob sie vielleicht von den Lehrkräften als Teil der täglichen Informationsflut einfach nicht mehr wahrgenommen worden waren. Doch eines ist unübersehbar: Die Mittel für diese gutgemeinte Maßnahme hätten ohne besondere Auswirkungen gut eingespart werden können.

Positiv erscheint in diesem Zusammenhang aber, dass ein Viertel der Befragten angab, über informelle Gespräche an der Schule von TIMSS erfahren zu haben. Dieser Befund wirft ein gutes Licht auf die informelle intraschulische Kommunikation und lässt vermuten, dass Lehrerinnen und Lehrer an den Schulen auch Fachgespräche führen, die über die unmittelbare, tagtägliche Arbeit hinausgehen. Die Tatsache wiederum, dass die schulinterne Lehrerfortbildung in diesem Zusammenhang nur eine nachrangige Bedeutung hatte, ist wenig erfreulich - aber für ein Land wie Baden-Württemberg, in dem es keine Verpflichtung zur internen Evaluation und zur Schulprogrammarbeit gibt, vielleicht auch nicht besonders unerwartet. Die Lehrerkonferenzen sind bekanntlich in der Regel mit Informationen, Diskussionen und Abstim-

mungen über organisatorische Dinge überfrachtet und werden kaum als Ort inhaltlicher Auseinandersetzung gesehen. Der zumeist einmal jährlich stattfindende pädagogische Tag wiederum bezieht sich - sinnvollerweise - nur auf ein einziges Thema, das gerade vor Ort vordringlich erscheint. Die Möglichkeit, im Kollegium kontinuierlich inhaltlich zu arbeiten, wird also bislang, allen Unterschieden zwischen den einzelnen Schulen zum Trotz, eher selten genutzt. Und doch lehrt uns die Schulentwicklungsforschung, dass für die Entwicklung und Sicherung der Qualität des Bildungswesens die Arbeit der einzelnen Schule entscheidend ist und sie *den* Ort für die Entwicklung bzw. Umsetzung von Innovationen darstellt (Peek, 2001; Rolff, 2001).

Betrachtet man sich die subjektive und die tatsächlich erhobene Informiertheit der Lehrkräfte an den einzelnen Schularten, so wird der geringe Informationsstand der Lehrerinnen und Lehrer an Grund- und Hauptschulen deutlich. Es wäre jedoch voreilig, aufgrund der vorliegenden Daten den Grund- und Hauptschullehrkräften ein allgemein geringes Interesse an pädagogischen und didaktischen Fragen zu unterstellen. Durchaus naheliegend erscheint es nämlich auch, dass sie sich weniger als Fachlehrerinnen und Fachlehrer verstehen, sie auch mehr mit allgemein erzieherischen Fragen befasst sind und sich von daher für primär fachdidaktisch erscheinende Themen weniger interessieren. Oder, um es etwas schroffer auszudrücken: Hauptschullehrerinnen und -lehrer sind oft schon froh, wenn sie eine Schulstunde mit vielleicht 32 Schülerinnen und Schülern ohne größere Konflikte überstehen und durchstehen - die Frage nach den fachlichen Leistungen im internationalen Vergleich erscheint da schnell sekundär. Grundschullehrerinnen und -lehrer schließlich fühlen sich von Untersuchungsergebnissen aus der Sekundarstufe vielleicht wenig berührt und erkennen so die Relevanz der Ergebnisse für ihre eigene Arbeit nicht. Auch wurde und wird die TIMS-Studie in Mitteilungen für die Primarstufe vergleichsweise selten zitiert.

Interessant erscheint der Befund, wonach die befragten Lehrerinnen und Lehrer die Ergebnisse der TIMS-Studie für nicht besonders bedeutsam hielten. Eine solche Einschätzung verträgt sich kaum mit den Reaktionen zahlreicher Lehrerbildungs- und -fortbildungsinstitutionen sowie weiterer wissenschaftlicher Einrichtungen, der zuständigen Ministerien oder auch der sogenannten Abnehmer. Ihre Initiativen oder auch nur verbalen Forderungen erscheinen nicht mit dem Rating der Lehrerinnen und Lehrer vereinbar, das nahe am theoretischen Mittelwert lag. Vermutlich messen Lehrerinnen und Lehrer den Ergebnissen der TIMS-Studie also weniger Bedeutung zu als dies andere mit Erziehung und Unterricht befasste Gruppen tun. Woran mag dies liegen? Nun, zunächst einmal ist davon auszugehen - und zeigte sich auch in der Untersuchung - dass die Einschätzung der Bedeutsamkeit der Ergebnisse unter anderem damit zu tun hat, für wie zutreffend die Ergebnisse leistungsvergleichender Studien im Allgemeinen und der TIMS-Studie im Besonderen gehalten werden. Im Hinblick auf diese Frage zeigten sich die befragten Lehrkräfte zwar nicht besonders skeptisch, doch kann nicht übersehen werden, dass sie der TIMS-Studie nicht kritiklos gegenüberstanden - sich vielleicht kritischer zeigten als andere Gruppen, die über die Anlage der

TIMS-Studie besser informiert sind oder weniger Probleme damit haben, die Schulleistungen einer ganzen Nation gewissermaßen durch einige wenige Zahlenwerte repräsentiert zu sehen. Hinzu kommt, dass Lehrerinnen und Lehrer, die sich tagtäglich mit einer immensen Informationsflut bezüglich ihrer schulischen Arbeit konfrontiert sehen, die exzeptionelle Stellung der TIMS-Studie im Vergleich zu einzelnen, lokalen Studien oder zu persönlichen Meinungen mehr oder weniger kompetenter Leute eher schwer ermessen können. Schließlich hören sie ständig, wie schlecht es um die Fähigkeiten und Fertigkeiten der Schülerinnen und Schüler in den unterschiedlichsten Bereichen und um die Qualität von Schule und Unterricht bestellt sei. Außerdem ist für Lehrerinnen und Lehrer eine an anderen Schulen durchgeführte Studie zunächst einmal, schon rein räumlich, weit weg. Eine solche Studie enthält keine Informationen über die Leistungen der eigenen Klasse, gibt somit auch keinen Aufschluss über den eigenen Unterrichtserfolg. Auch sind die von außen gesetzten Standards in den Augen vieler Lehrkräfte vermutlich nicht unbedingt geeignet, um den Ertrag und Erfolg der eigenen Arbeit überhaupt zutreffend zu beschreiben. Ihren Erfolg messen Lehrerinnen und Lehrer an eigenen Kriterien, und dies zumeist schon seit vielen Jahren. Insofern müssen die Ergebnisse der TIMS-Studie aus ihrer Sicht auch nicht unbedingt als besonders bedenklich eingestuft werden. Im Gegenteil, manche Lehrerinnen und Lehrer reagieren auf die Nachricht von den schwachen Leistungen der deutschen Schülerinnen und Schüler, gerade im Vergleich zu jenen der japanischen Lernenden, vielleicht mit wenig Betroffenheit: Da sie Drill und Druck oftmals als die zentralen Ursachenfaktoren des japanischen Erfolgs ausmachen - und die Erfolge anderer Nationen vernachlässigen -, besteht aus ihrer Sicht wenig Anlass, Konsequenzen aus der TIMS-Studie zu ziehen. Auch erhebt sich für viele Lehrerinnen und Lehrer mit ihrer Beurteilungskompetenz recht schnell die Frage, wozu es überhaupt einer vergleichenden Leistungsstudie bedarf. Diese Frage gewinnt dadurch an Brisanz, dass eine solche Studie natürlich immense Kosten verursacht und damit Ressourcen bindet, die aus Sicht der Lehrerinnen und Lehrer vielleicht an anderer Stelle viel nötiger gebraucht würden.

Die Tatsache, dass Grund- und Hauptschullehrkräfte die Ergebnisse der TIMS-Studie für etwas zutreffender hielten als ihre Kolleginnen und Kollegen an den beiden anderen Schularten, hat vermutlich mit den schwächeren Leistungen ihrer Schülerschaft zu tun, so dass sie die Hauptbotschaft der TIMS-Studie eher zu teilen vermochten. Die Lehrerinnen und Lehrer an Realschulen und erst recht an Gymnasien erlebten dagegen vielleicht bei Schüleraustauschprogrammen mit den USA, dass ihre Schülerinnen und Schüler den nordamerikanischen bezüglich bestimmter Kompetenzen überlegen waren, im internationalen Vergleich also gar nicht so schlecht da standen - und vergaßen dabei, dass sie die Leistungen ihrer selektierten Schülerschaft mit jenen von Schülerinnen und Schülern einer Gesamtschule verglichen.

Ein erfreuliches Ergebnis ist die eher gering ausgeprägte Skepsis gegenüber vergleichenden Leistungsstudien und generell gegenüber der wissenschaftli-

chen Arbeit an Hochschulen. Es lässt auf eine tendenziell vorurteilsarme und dann auch erfolgreiche Zusammenarbeit bei weiteren, noch ausstehenden Studien hoffen. Wichtig erscheint der Befund, dass Lehrerinnen und Lehrer in diesem Zusammenhang mit vergleichenden Leistungsstudien internationaler Prägung die größte Schwierigkeit haben. Daraus folgt, dass der Punkt des internationalen Vergleichs und vermutlich insbesondere der curricularen Validität der Testaufgaben für die einzelnen teilnehmenden Staaten sorgfältig und achtsam zu kommunizieren ist. Schließlich steht und fällt die Befürwortung und Unterstützung einer Studie mit der Frage, für wie valide die verwendeten Tests gehalten werden (Bos & Postlethwaite, 2001). Auch dürfte es in diesem Zusammenhang darauf ankommen, differenziert und verständlich zu vermitteln, welche Schlussfolgerungen die jeweilige Studie überhaupt zulässt - bzw. nicht zulässt. Eine verstärkte Einbindung von Lehrerinnen und Lehrern in die Konzeption externer Evaluationsstudien dürfte darüber hinaus eine überlegenswerte Maßnahme sein, um ein Mehr an Interesse, Aufgeschlossenheit und letztlich an Wirkung für die unterrichtliche Qualitätsentwicklung zu erzeugen (Rolff, 2001).

Nicht zu übersehen ist an dieser Stelle natürlich, dass die besondere Skepsis gegenüber vergleichenden Leistungsstudien internationaler Prägung im Anschluss an die TIMS-Studie mit ihren wenig erfreulichen Ergebnissen auch im Sinne der Attributionstheorie interpretiert werden kann. Doch der Hinweis auf die Notwendigkeit einer sorgsam Darstellung der Konzeption externer Evaluationsstudien erscheint auch unter diesem Blickwinkel gültig.

Bei der Frage nach den möglichen Ursachen des wenig erfreulichen Abschneidens der deutschen Schülerinnen und Schüler bei TIMSS, zu der bislang viele Überlegungen, aber nur wenige gesicherte Erkenntnisse vorliegen, ergab sich erwartungsgemäß ein Antwortverhalten, das Attributionsvoreingenommenheiten bei den befragten Lehrerinnen und Lehrern erkennen lässt. Die geringste Zustimmung erhielt die Skala „Ursache Lehrerinnen und Lehrer“, die einen Komplex internaler Ursachenfaktoren repräsentiert - verstärkt noch bei jenen Lehrerinnen und Lehrern, die Fachlehrkräfte für Mathematik und/oder Naturwissenschaften waren. Jene Skalen dagegen, die Beschreibungen externaler Ursachen umfassen, bekamen mit einer noch zu diskutierenden Ausnahme deutlich höhere Ratings. Interpretieren lassen sich diese typischen Attributionsvoreingenommenheiten einerseits als motivational bedingte Attributionsfehler, die im Falle von Misserfolg aus hedonistischen Gründen zu vornehmlich externalen Ursachenzuschreibungen führen, und andererseits als Ergebnisse unzureichender bzw. fehlerhafter Informationsverarbeitung bei den Lehrerinnen und Lehrern, die eine mangelnde Kovariation von Handlung und Ergebnis wahrnahmen: Da sie prinzipiell davon ausgingen, dass die Vertreterinnen und Vertreter ihres eigenen Berufsstandes sich anstrengen, die Ergebnisse der TIMS-Studie dies aber nicht unbedingt erkennen ließen, blieb ihnen bei der Erklärung von Misserfolg als logische Folge nur die Zuschreibung zu externalen Faktoren (Heckhausen, 1989; Weiner, 1994). Das Problem bei den Lehrerinnen und Lehrern dürfte hierbei unter anderem darin bestanden haben, dass sie vor dem Hintergrund ihrer ei-

genen Auffassung, was ein guter Unterricht sei, nachvollziehbar argumentierten, dabei aber verkannten, dass mit den Zielen und Methoden des bei uns üblichen, traditionellen Unterrichts ein wirklich gutes Abschneiden bei einer Studie wie TIMSS schwerlich möglich ist. Lehrerinnen und Lehrer arbeiten tagtäglich unter zum Teil immens schwierigen Bedingungen auf einen Unterricht hin, der, allen partiellen Modernisierungen zum Trotz, im Wesentlichen auf isoliertes Faktenwissen abzielt und die traditionelle Leistungsorientierung zumeist nicht verlässt (Gruber, Mandl & Renkl, 2000; Kohler, 2000). Die Tatsache, dass die Ergebnisse der TIMS-Studie uns auffordern, darüber nachzudenken, wie eigentlich ein guter Unterricht im begonnenen Jahrhundert aussehen könnte, über welche Kompetenzen unsere Schülerinnen und Schüler am Ende ihrer Schulzeit verfügen sollten und was uns hierzu andere Länder lehren können (Schrader & Helmke, 2001), wurde von den Lehrkräften unter dem Eindruck der täglichen Probleme, eingebunden in das gemeinsame, von der gesamten Umgebung geteilte Bezugssystem, vermutlich nicht reflektiert. Sie erleben aber im täglichen Unterricht immer wieder von neuem, wie sie ihre Ziele nicht erreichen können, weil ihre Schülerinnen und Schüler sich nicht in der von ihnen intendierten Weise verhalten. Insofern ist es kein Wunder, dass die Skala mit der höchsten Zustimmung die Skala „Ursache Schülerinnen und Schüler“ darstellt. Schaut man sich die Einzelitems mit den höchsten Ratings an, so sieht man diese Einschätzung noch einmal bestätigt: Die Lehrerinnen und Lehrer waren in hohem Maße der Meinung, die Schülerinnen und Schüler seien zu wenig leistungsbereit und wiesen eine zu geringe Konzentrationsfähigkeit auf (vgl. Tabelle 6). Sie schoben die Verantwortung für diese Defizite aber vermutlich weniger den einzelnen Schülerinnen und Schülern und deren Familien zu, sondern mehr den derzeitigen gesamtgesellschaftlichen Zuständen, die sie kritisch betrachteten. Damit nahmen sie eine externale Ursachenzuschreibung vor, die in unserer Gesellschaft vermutlich allgemein geteilt wird, somit keiner besonderen Begründung bedarf und auch wenig konflikthaft erscheint.

Der Blick auf den Unterricht selbst, der als Folge von TIMSS in der Forschung wieder in das Zentrum des Interesses gerückt ist (Baumert, Bos & Lehmann, 2000; Bos & Postlethwaite, 2001), spielte für die Lehrerinnen und Lehrer bei der Suche nach möglichen Erklärungen für das nur mittelmäßige Abschneiden der deutschen Schülerinnen und Schüler keine zentrale Rolle. Daraus dürfte resultieren, dass in der Schule und in der Forschung tätige Menschen im Hinblick auf die Ergebnisse von TIMSS - und sicher nicht nur da - ihren jeweils eigenen Fokus haben. Möchten also Erziehungswissenschaftlerinnen und -wissenschaftler die für sie zentralen Forschungsergebnisse und die dazugehörenden Implikationen Lehrkräften vermitteln, so können sie nicht unbedingt mit einem ihnen ausreichend erscheinenden Interesse rechnen. Es ist durchaus möglich, dass sie bei den Lehrerinnen und Lehrern zuallererst ein Bewusstsein für die Relevanz ihrer Überlegungen schaffen müssen. Ansonsten besteht die Gefahr, dass sich einerseits die Lehrkräfte mit ihren Fragen und Problemen von der Forschung im Stich gelassen fühlen und sich andererseits die wissenschaftlich Tätigen mit einem mangelnden

Interesse vonseiten der Lehrenden oder dem sogenannten Hochmut der Praktiker konfrontiert sehen.

Die Tatsache, dass Gymnasiallehrerinnen und -lehrer die Ursachen der nur mäßigen TIMSS-Testergebnisse im Vergleich zu ihren Kolleginnen und Kollegen an den beiden anderen Schularten in nicht so starkem Maße bei den Lernenden und ihren Eltern sahen, erstaunt sicher nicht, sondern erscheint mit Blick auf die Zusammensetzung der Schülerschaft bzw. deren soziale Herkunft leicht erklärlich. Die nicht vorhandene Differenz zwischen den Lehrkräften an Realschulen und jenen an Grund- und Hauptschulen lässt sich dagegen schwerer interpretieren. Naheliegend erscheint aber, dass die Lehrenden an Grund- und Hauptschulen ihre Eindrücke sowohl aus der Primarstufe (Gesamtschule) als auch aus der Sekundarstufe (Hauptschule) in ihr Rating einfließen ließen und so zu einer der Realschule entsprechenden Einschätzung gelangten.

Die geschilderten Attributionsvoreingenommenheiten der Lehrerinnen und Lehrer ziehen selbstverständlich Konsequenzen nach sich (Heckhausen, 1989; Weiner, 1994). Insbesondere ist davon auszugehen, dass die bevorzugt externalen Ursachenzuschreibungen nicht nur entlastend wirken, sondern - wie es auch die Datenlage nahelegt - außerdem zu einer geringeren Bereitschaft führen, sich selbst um eine Veränderung zu bemühen. Unter der Annahme, dass andere von TIMSS betroffene Gruppen an dieser Stelle ein anderes Antwortmuster aufweisen, ist weiterhin davon auszugehen, dass die Voreingenommenheiten der Lehrerinnen und Lehrer eine Verständigung unterschiedlicher Gruppen über die möglichen Implikationen von TIMSS erschweren. Zu bedenken ist an dieser Stelle aber auch, dass Lehrerinnen und Lehrer - wie kaum eine andere Berufsgruppe - ständig einer Art Kollektivschelte ausgesetzt sind, die dann über die Jahre dazu führen kann, dass eine Auseinandersetzung mit den eigenen Schwächen von vornherein abgelehnt wird. Die Tatsache, dass jüngere Lehrerinnen und Lehrer der Skala „Ursache Lehrerinnen und Lehrer“ mehr als ihre ältere Kolleginnen und Kollegen zustimmten, kann diese Vermutung stützen. Möglich im Hinblick auf diesen Punkt ist aber auch, dass jüngere Lehrerinnen und Lehrer beim Nachdenken über Schule und Unterricht über zweierlei Bezugssysteme verfügen - über ein an der Hochschule und ein an der Schule erworbenes -, wodurch sie zu einer kritischeren Haltung gegenüber der Arbeit an der Schule veranlasst werden. Die auf den ersten Blick vielleicht plausibel erscheinende Möglichkeit, dass jüngere Lehrerinnen und Lehrer sich noch nicht so stark mit ihrer Berufsgruppe identifizieren und ihre stärkere Zustimmung zur Skala „Ursache Lehrerinnen und Lehrer“ somit gar nicht im Sinne einer internalen Ursachenzuschreibung interpretiert werden darf, erscheint dagegen weniger wahrscheinlich: Es zeigte sich bei der Untersuchung, dass es keinen Zusammenhang zwischen dem Alter der Befragten und ihrer Einschätzung gab, ob sich ihr eigener Unterricht bzw. die Leistungen ihrer Schülerinnen und Schüler positiv von jenen unterscheiden würden, die bei TIMSS ermittelt worden waren.

Auf den ersten Blick erstaunlich mag erscheinen, dass Lehrkräfte mit Zusatzqualifikationen und/oder Zusatzfunktionen sich bei der Frage nach den Ursachen des nur mittelmäßigen Abschneidens der deutschen Schülerinnen und Schüler bei TIMSS fast nicht von ihren übrigen Kolleginnen und Kollegen unterscheiden. Schließlich liegt der Gedanke nahe, dass diese Lehrkräfte nicht nur über mehr bzw. differenzierteres fachliches und erzieherisches Wissen als ihre Kolleginnen und Kollegen verfügen, sondern auch eher um die Schwächen ihres Berufsstandes wissen, über dessen Durchschnitt sie sich zudem in irgendeiner Weise erheben wollten. Vielleicht ist es aber auch so, dass diese Personen in der Regel in ganz besonderer Weise in das Schulsystem eingebunden sind und somit gerade nicht über eine - kritische - Außensicht auf das System Schule verfügen. Sie geben als Mentoren, Fortbildner oder Schulleiterinnen vermutlich genau die überkommenen Standards eines guten Unterrichts weiter und können so nicht bemerken, dass mit den vorhandenen Zielen und vielleicht leicht modernisierten Methoden bei unseren Schülerinnen und Schülern die fehlende Problemlösekompetenz schwerlich aufgebaut werden kann. Die Tatsache, dass sich Lehrkräfte mit Zusatzqualifikationen und/oder Zusatzfunktionen auch in ihrem Vorwissen bezüglich TIMSS nicht bedeutsam von ihren Kolleginnen und Kollegen unterscheiden, deutet außerdem darauf hin, dass beispielsweise Fortbildner oder Mentorinnen neuere Themen der Forschung - aus ganz unterschiedlichen Gründen - auch nicht unbedingt besonders intensiv rezipieren und der Wissenstransfer von den Zentren der Lehreraus- und -fortbildung zur Basis nur schleppend oder wenig erfolgreich verläuft. Die an und für sich unstrittige Feststellung, in „... der Aus- und Weiterbildung von Lehrern haben die Fragestellungen und Befunde von TIMSS starken Widerhall gefunden“ (Baumert, Bos & Lehmann, 2000, S. 17), muss somit mit Blick auf die direkt an der Basis tätigen Mitarbeiterinnen und Mitarbeiter der Lehreraus- und -fortbildung ein Stück weit eingeschränkt werden.

Zunächst erstaunlich mag auch erscheinen, dass die Skala „Ursache Schulverwaltung und Bildungspolitik“ eine eher geringe Zustimmung erfuhr. Doch diese zunächst positiv erscheinende Einschätzung kann aufgrund der vergleichsweise hohen Zustimmung zu jenen Items, die sich auf die als unzureichend empfundene Unterstützung „von oben“ und auf die Bildungspolitik der letzten Jahre beziehen, nicht unbedingt als Resultat einer als positiv empfundenen Arbeit von Schuladministration und Bildungspolitik gedeutet werden. Möglich erscheint aber, dass die eher niedrigen Ratings ein Indiz dafür sind, dass sich Lehrerinnen und Lehrer an den Schulen tatsächlich als Teil des gesamten Systems begreifen und sie von daher die Grundlagen ihrer Arbeit, also z. B. Lehrpläne oder Schulbücher, wenig kritisieren, um sich und ihre Arbeit nicht selbst in Frage zu stellen. Diese Deutung erfährt auch Unterstützung durch die Tatsache, dass die Lehrerinnen und Lehrer die Qualität ihrer eigenen Ausbildung kaum in Frage stellen - und damit jenen Faktor, der in der Unterrichtsforschung als hoch bedeutsam gehandelt wird (z. B. Blum, 2000), nahezu negieren.

Interessant erscheint schließlich der Befund, wonach es für die Zustimmung zu den einzelnen Skalen keine Rolle spielte, ob die Lehrerinnen und Lehrer selbst Eltern waren. Dies kann - unter der Annahme, dass Eltern ein anderes Antwortmuster als Lehrkräfte aufweisen - dahingehend gedeutet werden, dass sich die Lehrerinnen und Lehrer, zumindest in dieser Befragung, primär als Lehrkräfte und nicht als Eltern begriffen.

Die Tatsache, dass die Befragten die Möglichkeit nicht ausschlossen, die baden-württembergischen Schülerinnen und Schüler im Allgemeinen und die eigenen Klassen im Besonderen hätten bei TIMSS bessere Ergebnisse als der Bundesdurchschnitt erzielt, lässt verschiedene Interpretationen zu. Zunächst einmal ist daran zu denken, dass allgemein davon ausgegangen wird, die Bundesländer Bayern und eben auch Baden-Württemberg setzten besonders hohe schulische Standards und die dort ansässigen Lernenden erbrächten entsprechend hohe Leistungen. Insofern erscheint es nicht weiters erstaunlich, dass auch die Befragten sich dieser Ansicht nicht verschlossen. Die Tatsache, dass die Zustimmung zu dieser Ansicht aber nicht besonders hoch ausfiel, könnte darauf hindeuten, dass die Hauptbotschaft der TIMS-Studie als durchaus korrekt, die TIMS-Studie somit als aussagekräftig auch für das eigene Bundesland eingeschätzt wurde. Besonders interessant erschien in diesem Zusammenhang die Frage, ob die Lehrerinnen und Lehrer den Ergebnissen der TIMS-Studie auch eine - unerfreuliche - Aussage über den eigenen Unterricht bzw. die Leistungen ihrer eigenen Schülerinnen und Schüler entnehmen würden oder ob sie sich von dieser Möglichkeit distanzieren. Die ermittelte Einschätzung, die nahe am theoretischen Mittelwert lag, lässt wiederum unterschiedliche Überlegungen zu. Zunächst einmal kommt man nicht umhin festzustellen, dass die befragten Lehrerinnen und Lehrer ihre eigene Arbeit bzw. die Leistungen ihrer Lernenden nicht ganz auf eine Stufe mit den bei TIMSS ermittelten Ergebnissen stellen wollten. Dann ist aber auch unübersehbar, dass die Befragten ihre Arbeit nicht als beträchtlich besser einschätzten als jene der Kolleginnen und Kollegen, die an TIMSS mitgewirkt hatten. Dies deutet darauf hin, dass sich die Befragten nicht grundsätzlich über andere Angehörige der eigenen Berufsgruppe stellen mochten und sich mit ihrer Berufsgruppe zumindest weitgehend identifizierten. Somit erscheint es tatsächlich erlaubt, bei den zuvor berichteten Attributionsvoreingenommenheiten im Falle der Zustimmung zur Skala „Ursache Lehrerinnen und Lehrer“ von internalen Ursachenzuschreibungen zu sprechen.

Auf den ersten Blick erfreulich und vielversprechend erscheint die insgesamt gesehen hohe Bereitschaft zur Veränderung und Verbesserung des eigenen Unterrichts, die sich insbesondere in einem hohen Interesse an entsprechenden Fortbildungsmaßnahmen und auch an geeignetem neuen Unterrichtsmaterial zeigte. Bei einem zweiten Blick kommt man jedoch nicht umhin festzustellen, dass es hier nur um ein unverbindlich geäußertes Interesse ging, das vielleicht auch aus Gründen sozialer Erwünschtheit so hoch ausfiel. Zudem ist zu bedenken, dass Lehrerinnen und Lehrer bei anderen Themen vielleicht ein mindestens ebenso großes Interesse bekundet hätten und es außerdem eines sehr großen Interesses bedarf, bis Lehrerinnen und Lehrer

sich bei ihrer zum Teil ausgesprochen hohen Arbeitsbelastung tatsächlich für eine Fortbildung entscheiden. Und dann kann an dieser Stelle natürlich nicht übersehen werden, dass eine vielleicht zweitägige Fortbildung oder die Einführung eines neuen Schulbuchs allein, also Maßnahmen, wie sie sich die Lehrkräfte vielleicht vorstellen, kaum bedeutsame Veränderungen langfristiger Art bewirken können. Erziehungs- und Lehrstile sowie pädagogisch-didaktische Kompetenzen von Lehrerinnen und Lehrern und erst recht die Erfolge ganzer Schulen sind das kumulative Ergebnis sehr langfristiger Prozesse und lassen sich durch singuläre und kurzfristige Maßnahmen nur geringfügig verändern (Weinert, 2001a). Wer wirklich eine Verbesserung der Leistungen der Schülerinnen und Schüler im Hinblick auf die Ergebnisse von TIMSS erreichen will, kommt, wie bereits weiter oben angeführt, an einer ganz grundsätzlichen Diskussion nicht nur um die Methoden, sondern auch um die Ziele unseres Schulunterrichts nicht herum.

Die Tatsache, dass das Alter der Befragten nur einen schwachen linearen Zusammenhang mit ihrer Bereitschaft zur Veränderung und Verbesserung des eigenen Unterrichts aufwies, lässt sich im Hinblick auf die älteren Lehrkräfte zunächst einmal positiv interpretieren. Dann ist jedoch auch an dieser Stelle wieder zu bedenken, dass lediglich eine Bereitschaft in unverbindlicher Weise abgefragt wurde und es nicht um tatsächlich durchgeführte Maßnahmen ging. Der beachtliche Zusammenhang zwischen Veränderungsbereitschaft und der jeweiligen Haltung zu kulturvergleichenden Studien schließlich erscheint gut nachvollziehbar: Wer die Seriosität und Aussagekraft vergleichender Studien generell in Frage stellt, braucht aus den dort gewonnenen Erkenntnissen auch keine Konsequenzen für sich selbst zu ziehen, sieht somit auch keine Notwendigkeit, sich um eine Veränderung seines Unterrichts Gedanken zu machen (Bethge, 1999).

Wirklich positiv erscheint die Tatsache, dass die befragten Lehrerinnen und Lehrer eine hohe Bereitschaft zur Teilnahme an - anonym durchgeführten - externen Evaluationsstudien bekundeten, letztere also offensichtlich nicht oder nicht ausschließlich - wie immer wieder formuliert - als Bedrohung oder Störung begriffen (Peek, 2001; Rolff, 2001). Dieses Ergebnis lässt zusammen mit der bereits erwähnten aufgeschlossenen Haltung gegenüber Untersuchungen und auch gegenüber den durchführenden Wissenschaftlern auf eine konfliktarme und erfolgreiche Durchführung weiterer Studien hoffen. Es fällt positiver aus als jenes einer Untersuchung des Instituts für Schulentwicklungsforschung der Universität Dortmund, bei der unter einer etwas anderen Fragestellung nur 29 % der Lehrerinnen und Lehrer für Fremdevaluation vonseiten der Schulaufsicht votierten (vgl. Rolff, 2001).

Klar erkennbar bei dieser Fragestellung ist die Bedeutung, die der Bedingung „Anonymität“ aus der Sicht der Lehrerinnen und Lehrern zukommt. Das hohe Interesse der Befragten am Abschneiden der eigenen Schülerinnen und Schüler weist darüber hinaus auf einen zweiten wichtigen Punkt hin. Lehrerinnen und Lehrer wollen vermutlich nicht einfach für irgendeine Institution Daten erbringen, und sie wollen ihre Arbeit auch nicht einfach „beur-

teilt“ und später in der Tageszeitung undifferenziert kritisiert sehen. Lehrende wünschen offensichtlich eine Rückmeldung, die sie ernst nimmt und - sozusagen als Gegenleistung - persönlichen Informationsgewinn bedeutet. Vermutlich wissen viele von ihnen, wie auch die Unterrichtsforschung, um die mangelnde Aussagekraft ihrer Leistungsbeurteilungen, wenn man den klasseninternen Vergleich verlässt, und sind von daher an entsprechenden Vergleichsdaten interessiert. Dieser ausgeprägte Wunsch der Lehrerinnen und Lehrer nach einem individuellen Feedback kann trotz des damit verbundenen erheblichen Aufwands als ausgesprochen erfreulich bezeichnet werden - insbesondere vor dem Hintergrund der immer wieder formulierten Befürchtungen, Lehrerinnen und Lehrer sperrten sich gegenüber den Aussagen externer Evaluation (vgl. Peek 2001). - Welche Faktoren über die erwähnten hinaus im Hinblick auf die Teilnahme an einer Studie von den Lehrerinnen und Lehrern als wichtig erachtet werden, sollte eine offene Frage innerhalb der Befragung erbringen, deren Auswertung aber einem späteren Zeitpunkt vorbehalten ist.

Der Extremgruppenvergleich zwischen Personen mit geringer vs. hoher Skepsis gegenüber externer Evaluation kann noch einmal einige der genannten Zusammenhänge aufzeigen. Dabei fällt insbesondere auf, dass es bei der Untersuchung tatsächlich eine hoch skeptische Gruppe von Personen gab, die vergleichsweise wenig Interesse an Informationsgewinn zeigte, sich nicht gern in Frage stellen ließ und an einer Veränderung eher gering interessiert war: Diese Gruppe hielt - im Vergleich zur Gruppe mit geringer Skepsis gegenüber externer Evaluation - die Ergebnisse der TIMS-Studie für deutlich weniger bedeutsam, zutreffend und bedenklich, zeigte bei der Frage nach den Ursachen des Abscheidens der deutschen Schülerinnen und Schüler eine signifikant stärkere externale Ursachenzuschreibung, war außerdem viel eher der Meinung, die baden-württembergischen Schülerinnen und Schüler im Allgemeinen und die eigenen Klassen im Besonderen hätten besser als der Bundesdurchschnitt abgeschnitten, war deutlich weniger bereit, an einer vergleichenden Leistungsstudie teilzunehmen und hatte schließlich auch ein signifikant geringeres Interesse am Abscheiden der eigenen Schülerinnen und Schüler bei einer entsprechenden Studie einerseits und an einer Veränderung und Verbesserung des eigenen Unterrichts andererseits. Diese Gruppe von Lehrerinnen und Lehrern zeigte im Einzelnen zwar nicht unbedingt bedenkliche Ratings, verfügte aber über ein Antwortmuster, das auf eine gewisse Veränderungsresistenz hindeutet und im Falle der Durchführung leistungsvergleichender Studien die Kommunikation mit den jeweils Durchführenden erschweren dürfte.

Aufgabe einer weiteren Studie wird es daher sein, unter Nutzung der eingangs erwähnten generierbaren Codes mit einigen Personen der beiden Extremgruppen in Kontakt zu treten und so die jeweiligen Hintergründe der beschriebenen Antwortmuster zu beleuchten. Geplant ist außerdem die Auswertung einer vergleichenden Befragung von Eltern sowie von Beamtinnen und Beamten der Schulverwaltung. Mit Hilfe der dabei gewonnenen Daten sowie den Ergebnissen der hier beschriebenen Untersuchung soll dann unter

anderem der Frage nachgegangen werden, ob alle drei von den Ergebnissen der TIMS-Studie betroffenen Gruppen - Lehrende, Eltern und Beamtinnen und Beamte der Schuladministration - typische Attributionsvoreingenommenheiten erkennen lassen.

Literaturverzeichnis

- Baumert, J., Bos, W. & Lehmann, R. H. (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Bd. 1. Opladen: Leske + Budrich.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS. Mathematisch-naturwissenschaftlicher Unterricht im Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Behörde für Schule, Jugend und Berufsbildung Hamburg (Hrsg.) (2000). *Hinweise für die Auswertung des wissenschaftlichen Berichts und der klassenbezogenen Rückmeldungen. Aspekte der Lernausgangslage und der Lernentwicklung - Klassenstufe 7*. Anlage 2. Hamburg.
- Bethge, T. (1999). Zum Umgang mit den Ergebnissen von TIMSS. In *Die Deutsche Schule*, 91 (2), 178 - 181.
- Blum, W. (2000). Qualitätsentwicklung im Mathematikunterricht ... eine Folge von TIMSS? In *Pädagogik*, 52 (12), 23-26.
- Bos, W. & Postlethwaite, T. N. (2001). Internationale Schulleistungsforschung: Ihre Entwicklungen und Folgen für die deutsche Bildungslandschaft. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 251-268). Weinheim: Beltz.
- Demmer, M. (2000). Risiken und Nebenwirkungen von Schulleistungsvergleichen. Argumente gegen Testeuphorie. In *Pädagogik*, 52 (12), 32-35.
- Enzmann, D. & Kleiber, D. (1989). *Helfer-Leiden: Streß und Burnout in psychosozialen Berufen*. Heidelberg: Asanger.
- Gruber, H., Mandl, H. & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmaier (Hrsg.), *Die Kluft zwischen Wissen und Handeln*. Göttingen: Hogrefe.
- Heckhausen, H. (1989). *Motivation und Handeln*, 2. Auflage. Berlin: Springer.
- Helmke, A. & Schrader, F.-W. (2001). Jenseits von TIMSS: Messungen sprachlicher Kompetenzen, komplexe Längsschnittstudien und kulturvergleichende Analysen. Ergebnisse und Perspektiven ausgewählter Leistungsstudien. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 237-250). Weinheim: Beltz.
- Klieme, E., Baumert, J. & Schwippert, K. (2000). Schulbezogene Evaluation und Schulleistungsvergleiche. In H. G. Rolff (Hrsg.), *Jahrbuch der Schulentwicklung*. Bd. 11. Weinheim: Juventa.
- Klieme, E., Knoll, S. & Schümer, G. (1999). *Mathematikunterricht der Sekundarstufe I in Deutschland, Japan und den USA. Dokumentation zur TIMSS-Videostudie* [CD-ROM]. Stuttgart: Klett.
- Kohler, B. (2000). Problemlöseaufgaben bewältigen und Kenntnisse erwerben: Lernen mit problemorientiert gestalteten Texten. In *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 32 (1), 34-43.
- Peek, R. (2001). Die Bedeutung vergleichender Schulleistungsmessungen für die Qualitätskontrolle und Qualitätsentwicklung von Schulen und Schulsystemen. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 323-336). Weinheim: Beltz.

- Rolff, H.-G. (2001). Was bringt die vergleichende Leistungsmessung für die pädagogische Arbeit in Schulen? In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 337-352). Weinheim: Beltz.
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S.45-58). Weinheim: Beltz.
- Sygmund, H. (1999). Bildung und Leistung gehören zusammen! In *Die Deutsche Schule*, 91 (2), 182-185.
- Weiner, B. (1994). *Motivationspsychologie*, 3. Auflage. Weinheim: PVU.
- Weinert, F. E. (2001a). Perspektiven der Schulleistungsmessung - mehrperspektivisch betrachtet. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 353-366). Weinheim: Beltz.
- Weinert, F. E. (2001b). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17-32). Weinheim: Beltz.

Anschrift der Autorin:

Dr. Britta Kohler

Pädagogische Hochschule Schwäbisch Gmünd

Oberbettringer Straße 200

73525 Schwäbisch Gmünd

britta.kohler@t-online.de